Title

Implementation Intentions Reduce Implicit Stereotype Activation and Application.

Journal

Personality & social psychology bulletin, 45(1)

ISSN

0146-1672

Authors

Rees, Heather Rose
Rivers, Andrew Michael
Sherman, Jeffrey W

Publication Date

2019

DOI

10.1177/0146167218775695

Peer reviewed

Empirical Research Paper

Implementation Intentions Reduce Implicit Stereotype Activation and Application

Personality and Social Psychology Bulletin
1–17

DOI: 10.1177/0146167218775695
pspb.sagepub.com
SAGE

Heather Rose Rees[1], Andrew Michael Rivers[1], and Jeffrey W. Sherman[1]

Abstract

Research has found that implementation intentions, if–then action plans (e.g., "if I see a Black face, I will think safe"), reduce stereotyping on implicit measures. However, it is unknown by what process(es) implementation intentions reduce implicit stereotyping. The present research examines the effects of implementation intentions on stereotype activation (e.g., extent to which stereotypic information is accessible) and stereotype application (e.g., extent to which accessible stereotypes are applied in judgment). In addition, we assessed the efficiency of implementation intentions by manipulating cognitive resources (e.g., digit-span, restricted response window) while participants made judgments on an implicit stereotyping measure. Across four studies, implementation intentions reduced implicit stereotyping. This decrease in stereotyping was associated with reductions in both stereotype activation and application. In addition, these effects of implementation intentions were highly efficient and associated with reduced stereotyping even for groups for which people may have little practice inhibiting stereotypes (e.g., gender).

Keywords

implementation intentions, implicit stereotyping, stereotype activation, stereotype application, multinomial modeling

Received October 14, 2017; revision accepted April 16, 2018

For both personal and social reasons, people may be unwilling to report prejudices toward different social groups. Even when willing, people may be unaware of their biases and, therefore, unable to report them (Fazio, Jackson, Dunton, & Williams, 1995). Implicit measures of intergroup bias were developed to address these shortcomings by measuring bias without asking respondents to directly report them (Banaji & Greenwald, 1994; Greenwald & Banaji, 1995). Such bias[1] has turned out to be pervasive, and predictive of stereotyping behavior. For example, large samples of data taken from project implicit show that pro-White/anti-Black bias is common, with an average effect size of $d = .80$ (Lane, Banaji, Nosek, & Greenwald, 2007). In addition, implicitly measured race bias is correlated with meaningful behaviors, such as racial disparities in police shooting, $\beta = .39$ (Hehman, Flake, & Calanchini, 2017). Given its potential for adversely affecting behavior and outcomes, it is not surprising that researchers have taken a keen interest in understanding how such biases may be reduced. As such, in recent years, many interventions have been developed to decrease implicit bias (however, they may be short-lived; Dovidio, Kawakami, & Gaertner, 2000; Lai et al., 2014).

Many bias interventions directly target beliefs about a group. For example, one strategy is to ask participants to vividly imagine a counter-stereotypic scenario in which a Black individual behaves more positively than a White individual (Foroni & Mayr, 2005; Lai et al., 2014). Another approach is to directly present participants with counter-stereotypic Black and White exemplars (Dasgupta & Greenwald, 2001; Joy-Gaba & Nosek, 2010; Lai et al., 2014). The goal of these interventions is to either change the accessibility of different aspects of group knowledge or to directly alter, if only temporarily, the extent to which different attributes are associated with social groups. They are attempts to change the underlying mental representations that produce bias.

In contrast to interventions that focus on changing underlying beliefs about a group, other interventions attempt to reduce bias by providing people with strategies for how to respond without bias while completing the implicit measure (e.g., Fiedler & Bluemke, 2005; Lai et al., 2014). Perhaps the most promising of these interventions is to equip people with "implementation intentions" that offer concrete behavioral

[1]University of California, Davis, USA

Corresponding Author:
Heather Rose Rees, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA.
Email: hrrees@ucdavis.edu

plans for avoiding the expression of bias (Lai et al., 2014; Mendoza, Gollwitzer, & Amodio, 2010; Stewart & Payne, 2008). Implementation intentions are if–then action plans in which individuals form an association between a cue (the "if") and a desired behavior (the "then"). Implementation intentions have been used successfully to change both thoughts and behavior in several domains, including academic achievement, dieting, and many others (Gollwitzer & Sheeran, 2006). For example, if one wanted to think of Black people as less threatening, one could form an implementation intention: "If I see a Black person, I will think safe." Because the cue (a Black person) and the behavior (thinking safe) are associated and planned in advance, the behavior is more likely to occur than if an individual merely forms general intentions to think of Black individuals as safe (Brandstätter, Lengfelder, & Gollwitzer, 2001). As a result of forming such implementation intentions, the mere presentation of a cue (a Black person) can automatically facilitate the desired behavior (thinking safe). For example, Stewart and Payne (2008) found that participants who formed implementation intentions to think "safe" whenever they saw a Black face during the Weapon Identification Task (Payne, 2001) were less influenced by racial stereotypes compared to a control condition. Similarly, Mendoza et al. (2010) found that implementation intentions to ignore race decreased racial bias on the First-Person Shooter Task (Correll, Park, Judd, & Wittenbrink, 2002). Thus, implementation intentions have been shown to successfully reduce implicit intergroup bias and may even be able to do so for longer periods of time than other interventions (up to 3 weeks; Webb, Sheeran, & Pepper, 2012).

Given the promise of implementation intentions as an effective and lasting intervention to reduce implicit bias, it is vital to understand the mechanisms underlying such effects. To the extent that these mechanisms are understood, aspects of the intervention can be fine-tuned to maximize its effectiveness. Moreover, such knowledge may permit the development of new interventions that target the same mechanisms. Toward this end, researchers have attempted to delineate the components of implicit bias that are affected by implementation intentions. Although implicit bias is often described as reflecting only unintentional or automatic responses, there is now considerable evidence that intentional processes contribute substantially to the extent of implicit bias (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Krieglmeyer & Sherman, 2012; Payne, 2001). Stewart and Payne (2008) and Mendoza et al. (2010) used the process dissociation technique (e.g., Payne, 2001) to assess the extent to which implementation intentions affect intentional versus unintentional components of implicit bias. Whereas Stewart and Payne (2008) found that implementation intentions reduced unintended but not intended components of bias, Mendoza et al. (2010) found evidence that both unintentional and intentional processes were affected.

Although process dissociation separates the influence of intentional versus unintentional processes, it does not identify the contributions of specific cognitive mechanisms by which implementation intentions may induce their effects. Of particular interest is the extent to which implementation intentions reduce implicit bias by reducing the extent to which stereotypes are activated in memory (stereotype activation) versus reducing the extent to which activated stereotypes are applied during judgment processes (stereotype application; Devine & Monteith, 1999; Gilbert & Hixon, 1991; Kunda & Spencer, 2003). Theoretically, stereotype activation must precede stereotype application. Unless stereotypes are activated in the first place, there is no stereotype to apply or correct against. In part, because stereotype application comes later in the temporal sequence, most models of stereotyping assume that stereotype application is easier to control than is stereotype activation. That is, with stereotype application occurring later in the sequence, there is greater opportunity to impose one's intentions on stereotype application than activation (e.g., Fazio et al., 1995). As such, if implementation intentions reduce stereotype application, but not activation, then any effects on reduced bias may be relatively fragile. Specifically, if anything interferes with the ability to prevent application (e.g., limited cognitive resources; limited response time), then activated stereotypes will affect judgment and behavior. In contrast, if implementation intentions can reduce the initial activation of stereotypes, then such effects would presumably be more robust against interference. If the stereotype is not activated in the first place, it will not be applied, even under conditions of limited resources or time (Devine & Monteith, 1999; Gilbert & Hixon, 1991; Kunda & Spencer, 2003). Thus, for both theoretical and practical reasons, it is important to know whether implementation intentions reduce stereotype activation, stereotype application, or both.

The Current Research

The main goal of the current research was to test the extent to which implementation intentions influence stereotype activation and stereotype application. Specifically, we test whether implementation intentions formed to respond without bias do so by decreasing stereotype activation or stereotype application. The second goal of this research was to examine the extent to which these effects of implementation intentions depend on the availability of cognitive resources. We know that implementation intentions effectively reduce racial bias on implicit measures. This alone suggests that implementation intentions may operate relatively efficiently. Nevertheless, implementation intentions are an intentional strategy for reducing racial bias. Intentionality is thought to be affected by the extent to which individuals have access to cognitive resources. As such, it is possible that implementation intentions require sufficient cognitive resources to affect implicit bias (Conrey et al., 2005; Govorun & Payne, 2006). This would

set an important constraint on the contexts in which implementation intentions might be expected to be effective interventions. Thus, for the first time, we directly investigate the extent to which implementation intentions depend on cognitive resources to reduce implicit bias and the processes that contribute to bias (i.e., activation and application).

A related question concerns the role of experience or practice in inhibiting stereotypes. To the extent that people have practiced inhibiting the expression of a stereotype, the process should become relatively routinized and automatic. We explore this issue by examining the effects of implementation intentions on stereotypes that our participants attempt to inhibit more (race) or less (gender) frequently. In particular, among our college-aged participants, social norms often encourage inhibition of stereotypes of Black men as threatening. In contrast, there are no social norms dictating that men should not be judged as more threatening than women. As such, our participants should be more practiced at inhibiting race than gender stereotypes surrounding threat. If the effects of implementation intentions are restricted to highly practiced stereotype inhibition (see Moskowitz & Li, 2011 for data consistent with this possibility), this would suggest an important constraint on their effectiveness. In contrast, if implementation intentions are effective even with stereotypes that are not typically inhibited, then this would indicate that the strategy is broadly applicable for bias reduction.

Measuring Stereotype Activation and Application

To investigate the mechanisms by which implementation intentions work to reduce implicit racial bias, we examine their effects on performance on the stereotype misperception task (SMT; Krieglmeyer & Sherman, 2012). The SMT was developed, specifically, to measure both stereotype activation and application during performance of a single task. It is the only existing means to measure these processes simultaneously and independently. Typically, stereotype activation and application are measured via performance on different measures (for a review, see Krieglmeyer & Sherman, 2012). Stereotype activation is most often measured via some form of priming measure that is presumed to provide an implicit and relatively pure assessment of the extent to which knowledge has been activated in memory. In contrast, stereotype application is most often measured via performance on some judgment task, in which participants provide explicit judgments of a target based on different arrays of information. The underlying logic of this approach is that performance on measures of activation reflects only the extent of activation and not other stereotyping processes. In contrast, performance on measures of application have been presumed to reflect primarily intentional judgment processes that are engaged after stereotypes become active (e.g., whether to apply the stereotype or not). However, there is now extensive evidence that performance on measures of stereotype activation is influenced by intentional processes. In other words, measures of stereotype activation and application are not pure measures of the intended processes but, rather, reflect the influence of multiple processes (e.g., Krieglmeyer & Sherman, 2012; Payne, 2001).

More broadly, there is always the risk when using different tasks to measure different processes that any observed differences in performance may reflect specific procedural features of the measures rather than differences in the processes of interest (e.g., "structural fit"). Thus, when using priming versus judgment tasks to measure stereotype activation and application, respectively, it is always possible that any observed differences reflect a myriad of method-related processes that differ between the tasks, rather than differences in the extents of activation and application. Therefore, if one wants to test and compare the effects of a manipulation, such as implementation intentions, on two different processes (e.g., stereotype activation vs. stereotype application), it is desirable to keep the procedural features of the measures constant. One solution to this problem is to use a single task that reflects the joint contributions of both processes and apply modeling techniques (described below) to disentangle the processes of interest (for a review, see Sherman, Klauer, & Allen, 2010). The SMT was designed, specifically, to measure the extents of stereotype activation and application separately and independently from performance on a single task, thereby controlling for any method variance in the measurement of the processes. The ability of the SMT to accomplish this goal has been confirmed through careful validation studies (Krieglmeyer & Sherman, 2012).

Overview

In Experiment 1, we examine whether implementation intentions reduce stereotyping on the SMT, and the processes by which they do so. Specifically, we examine the extent to which implementation intentions reduce bias by influencing stereotype activation and application. In Experiment 2, we also investigate the efficiency with which implementation intentions operate by directly replicating Experiment 1 while manipulating the cognitive resources available to respondents. In Experiment 3, we further examine the efficiency of implementation intentions by reducing cognitive resources by manipulating the time participants have to make judgments. Finally, in Experiment 4, we examine whether the effectiveness and efficiency of implementation intentions extends to a stereotype that people are unlikely to regularly inhibit: the stereotype that men are more threatening than women. This provides further insight into the efficiency, robustness, and generality of implementation intention effects.

Experiment 1

Method

Preregistration. We preregistered each experiment in the current paper on the Open Science Framework (Open Science Collaboration, 2012; all preregistration materials available at

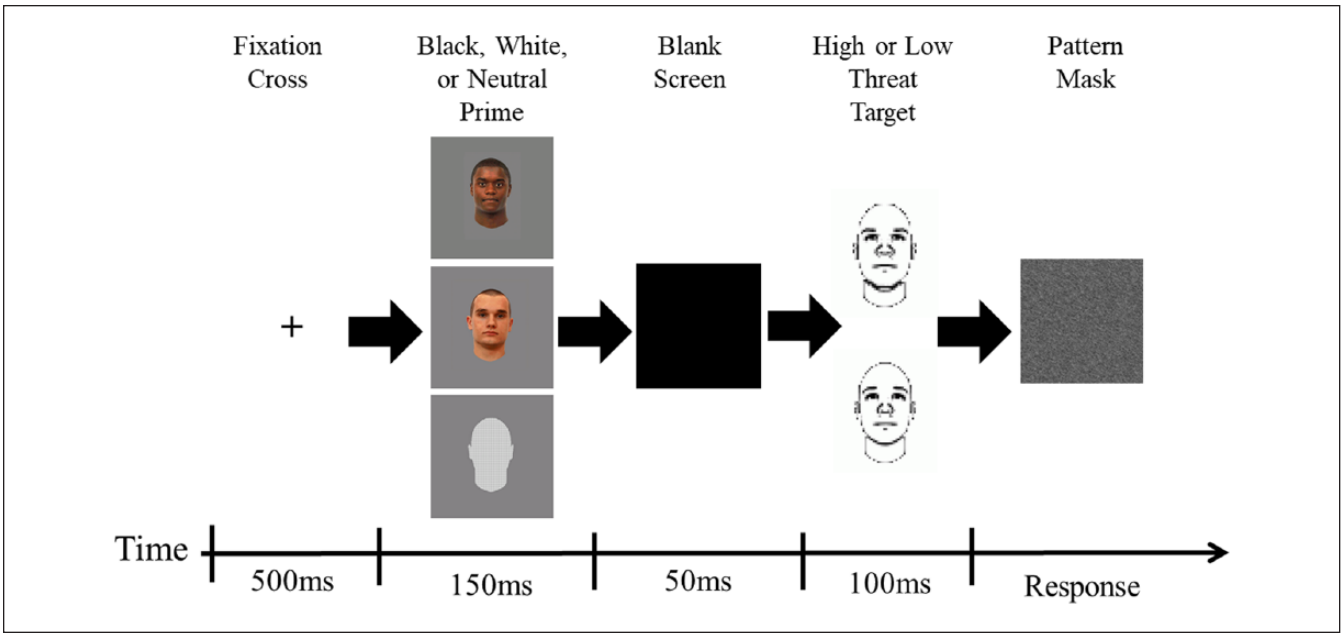

Figure 1. Visual representation of the SMT Procedure.
Note. SMT = stereotype misperception task.

https://osf.io/vhyex/). Preregistration plans included a priori power analyses and the minimum number of participants we sought to sample. The plan also included exclusion criteria, including excluding data from (a) participants making the same judgment on all SMT trials, (b) participants making the same judgment at a rate ±3 *SD* from their group, and (c) participants incorrectly reporting their condition assignment in a post-experiment manipulation check. We note for each experiment the number of participants excluded for these criteria. In addition, data were not analyzed at any intermediate point prior to obtaining the final sample in each experiment. We report all measures and conditions.

Participants. We planned to collect data from at least 123 participants, providing $1 - \beta = .80$ power, given the effect size ($\eta_p^2 = .09$) reported in Stewart and Payne (2008). In total, 237 participants from the University of California, Davis completed the experiment for partial course credit (providing greater than .95 power). In Experiment 1, eight participants made the same judgment on all trials and an additional 19 failed to report their condition assignment.[2] Finally, research assistants identified two additional cases of strange behavior during the experiment (i.e., falling asleep, not attending to the task).[3] The final sample consisted of 210 participants across three between-subjects conditions ("safe" $N = 83$, "accurate" $N = 63$, "quick" $N = 64$).

Stimuli. Prime stimuli were photographs of 24 Black and 24 White males that were cropped at the neck (adapted from Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). In addition to the racial stimuli, we also included a neutral prime image that consisted of a facial outline of the same shape and size as the other photographs (see Krieglmeyer & Sherman, 2012). All prime stimuli were superimposed on a gray background (see Figure 1).

Target stimuli were 48 computer-generated facial morphs developed by Oosterhof and Todorov (2008) that were degraded using a pixelation filter in Adobe Photoshop (see Krieglmeyer & Sherman, 2012). These stimuli were made from 24 unique identity faces that objectively varied in threat. That is, each unique identity had an associated face that was two standard deviations above (high threat) or below (low threat) a neutral point of threat ratings.

Procedure

SMT. Participants completed the experiment individually on computers in groups of 1 to 4. They learned that the task (SMT) tested their ability to form rapid impressions and were instructed to judge whether target faces were more or less threatening than the average target presented in the task. Participants were instructed to respond as quickly as possible and to rely on their "gut" impressions in judging target faces. They were told to attend to prime faces for later questions but were also told that they should avoid responding to the prime faces when judging the computer-generated target faces. Participants completed two blocks of six practice trials each to ensure that they understood the procedure (see Krieglmeyer & Sherman, 2012).

Participants then completed two test blocks of 72 trials each. Each test trial started with a fixation cross in the center of the screen for 500 ms, which was followed by a prime picture for 150 ms. After the prime picture, there was a blank

Table 1. Proportion of Threat Judgments (and Standard Errors) as a Function of Implementation Intention, Prime, and Target.

Experiment 1			
	Black	Neutral	White
Accurate			
High threat	.48 (.03)	.31 (.04)	.34 (.03)
Low threat	.44 (.03)	.24 (.03)	.33 (.03)
Quick			
High threat	.50 (.03)	.31 (.04)	.34 (.03)
Low threat	.47 (.03)	.23 (.03)	.34 (.03)
Safe			
High threat	.25 (.03)	.42 (.03)	.49 (.03)
Low threat	.23 (.03)	.33 (.03)	.46 (.03)

Note. Primes listed horizontally in the top row. Targets listed vertically in column.

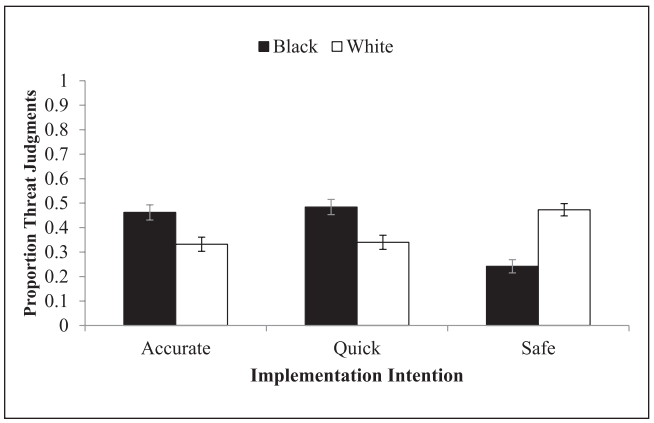

Figure 2. Proportion of threat judgments as a function of implementation intention and prime in Experiment 1.
Note. Error bars represent standard error of the mean.

screen for 50 ms and then a target image for 100 ms. Following the target image presentation, a gray pattern mask was displayed until participants made a key press response (selecting either "more" or "less" threatening). A 500 ms intertrial interval followed each response. All prime types were paired equally with each target type. Stereotypic biases on the SMT are evident if participants judge a greater proportion of targets as more threatening after Black than after White primes, despite explicit instructions to avoid being influenced by prime pictures.

Implementation intention instructions. Participants were randomly assigned to one of three implementation intention conditions, identical to those used in Stewart and Payne (2008). They were instructed to form implementation intentions to think "Safe," "Accurate," or "Quick," whenever they saw a Black face. After forming implementation intentions, participants were reminded to rely on their immediate gut feelings to make their judgments but were additionally told, "from now on, be sure to think (safe/accurate/quick) when you see Black faces." Following Stewart and Payne (2008), we expected the Accurate and Quick intention conditions to act as control conditions compared to the critical Safe condition, which was expected to reduce the extent of bias. We did not expect meaningful differences between the Accurate and Quick conditions, but, given that we were adapting a manipulation from prior research, we opted to be cautious and run both of the control conditions reported in previous work.

Design. The experiment had a 3 (Intention: Safe vs. Accurate vs. Quick) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target: High threat vs. Low threat) mixed design. Implementation intentions were manipulated as a between-subjects factor, whereas prime and target were within-subjects factors.

Results[4]

SMT effect. To test the effectiveness of implementation intentions on racial bias, we subjected the proportion of "more threatening" responses to a 3 (Implementation Intention: Accurate vs. Quick vs. Safe) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target Threat: High vs. Low) mixed analysis of variance (ANOVA). There was a main effect of target, $F(1, 207) = 32.29$, $p < .001$, $\eta_p^2 = .14$, such that high threat targets were given a greater proportion of threat judgments than low threat targets (see Table 1). There was also a prime main effect, $F(1.91, 395.14)$[5] $= 7.90$, $p = .001$, $\eta_p^2 = .04$. Simple comparisons showed that targets were judged as more threatening following Black primes compared to Neutral primes, $F(1, 209) = 6.07$, $p = .015$, $\eta_p^2 = .03$, but not compared to White primes, $F(1, 209) = .10$, $p = .758$. More threat judgments were given after White primes compared to Neutral primes, $F(1, 209) = 14.17$, $p < .001$, $\eta_p^2 = .06$.[6]

As expected, implementation intentions moderated the prime effect, $F(4, 414) = 16.91$, $p < .001$, $\eta_p^2 = .14$ (see Figure 2). To better understand this interaction, we conducted three separate repeated-measures ANOVAs on the proportion of "more threatening" responses to examine the prime effect within each intention condition.

A main effect of prime was evident in the Accurate condition, $F(2, 124) = 9.18$, $p < .001$, $\eta_p^2 = .13$. Simple comparisons indicated that a greater proportion of threat judgments were made following Black primes than White primes, $F(1, 62) = 7.89$, $p = .007$, $\eta_p^2 = .11$, 95% CIdiff [.04, .22], and Neutral primes, $F(1, 62) = 16.17$, $p < .001$, $\eta_p^2 = .21$, 95% CIdiff [.10, .28]. In the Quick intention condition, there also was a prime main effect, $F(1.85, 116.76) = 16.49$, $p < .001$, $\eta_p^2 = .21$. Simple comparisons showed that there was a higher proportion of threat judgments following Black primes than White primes, $F(1, 63) = 12.08$, $p = .001$, $\eta_p^2 = .16$, 95% CIdiff [.06, .23], and Neutral primes, $F(1, 63) = 27.70$,

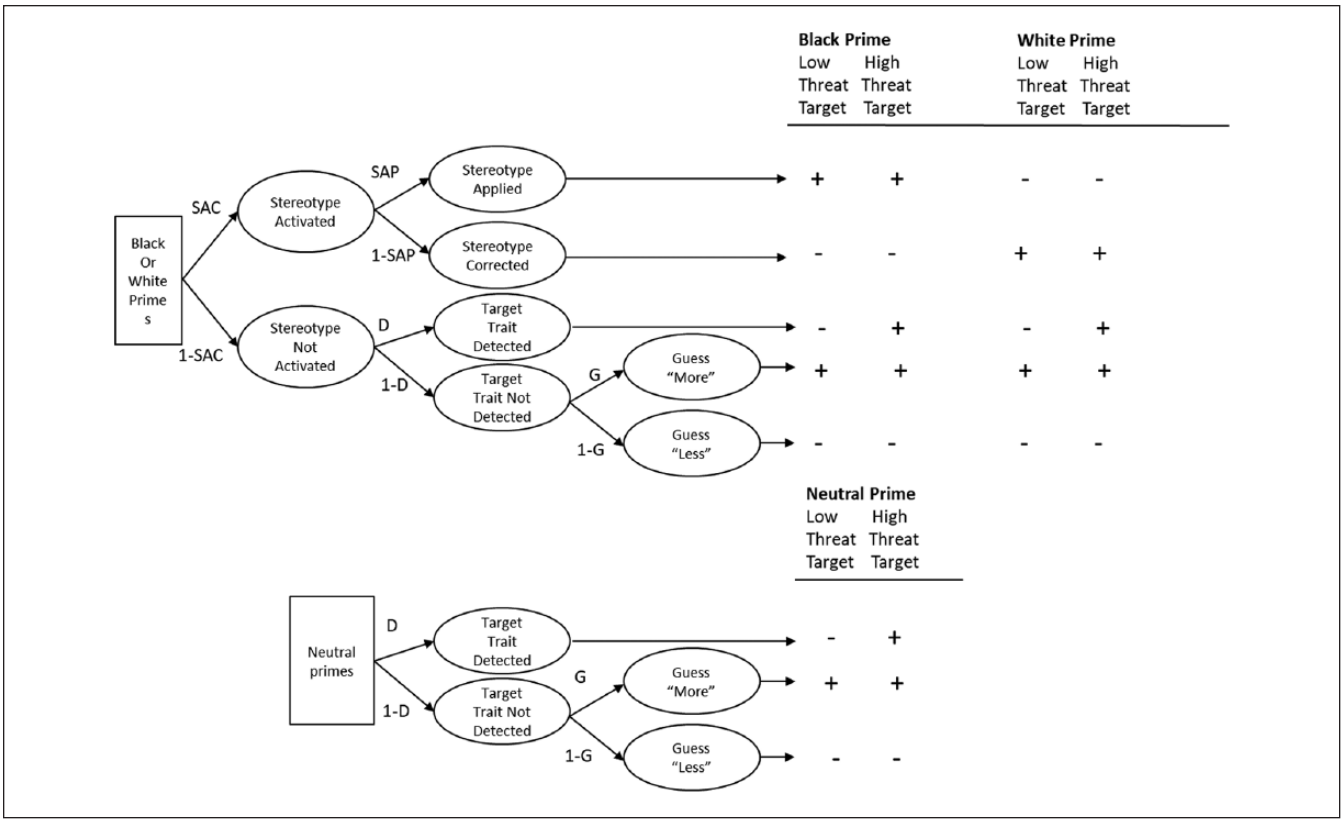

Figure 3. Structure of the SMT multinomial processing tree.
Note. The top part shows the model for Black and White primes, and the bottom part shows the model for neutral primes. The table on the right depicts the responses as a function of prime and target. The response "more threat" is represented by a + sign and the response "less threat" is represented by a − sign. SAC = stereotype activation; 1 − SAC = lack of stereotype activation; SAP = stereotype application; 1 − SAP = stereotype correction; D = detection of target threat; 1 − D = detection of target trait; G = tendency to guess "more threat"; 1 − G = tendency to guess "less threat."
SAC = stereotype activation; SAP = stereotype application; SMT = stereotype misperception task.

$p < .001$, $\eta_p^2 = .31$, 95% CIdiff [.13, .30]. There also was a greater proportion of threat judgments following White primes than Neutral primes, $F(1, 63) = 5.12$, $p = .027$, $\eta_p^2 = .08$, 95% CIdiff [.01, .13]. These results indicate that racial bias was statistically detectable in both the Accurate and Quick conditions; participants more often judged targets as threatening when following Black versus White primes.

In the Safe condition, there also was a prime effect, $F(1.90, 152.71) = 15.68$, $p < .001$, $\eta_p^2 = .16$. In contrast to the Accurate and Quick conditions, simple comparisons indicated that the proportion of threat judgments was *lower* for targets following Black primes than White primes, $F(1, 82) = 28.27$, $p < .001$, $\eta_p^2 = .26$, 95% CIdiff [−.32, −.15], and Neutral primes, $F(1, 82) = 8.46$, $p = .005$, $\eta_p^2 = .09$, 95% CIdiff [−.22, −.04]. Thus, the prime effect observed in the Accurate and Quick conditions was reversed in the Safe condition. The proportion of threat judgments was higher following White primes than Neutral primes, $F(1, 82) = 8.22$, $p = .005$, $\eta_p^2 = .09$, 95% CIdiff [.03, .17].

Multinomial modeling overview. To investigate the cognitive mechanisms responsible for the effectiveness of implementation intentions on the SMT, we applied the SMT multinomial processing tree model. Multinomial models, such as the SMT model, attempt to describe experimental outcomes (e.g., proportion of threat judgments) via a set of variables (or parameters) and a set of equations that establish relationships among the variables. The variables in the equations represent the hypothesized component processes (e.g., stereotype activation; stereotype application). Solving for these variables yields independent estimates of the extent of each process. To achieve the measurement of distinct processes, parameter estimates are systematically varied through maximum likelihood estimation to determine the values that most closely reproduce actual task performance. Application of the SMT model allows us to measure the extent of stereotype activation and application separately and independently from performance on the SMT.

The SMT model estimates four parameters: stereotype activation, stereotype application, detection of target threat (D), and guessing tendencies (G). The SMT model describes how all of the four processes interact to produce responses on the SMT (see Figure 3). Racial primes activate stereotypes with the probability (SAC) or fail to activate stereotypes with the probability of (1 − SAC). If a stereotype is activated, it is applied with the probability (SAP) or is not

applied in judgment with the probability of (1 − SAP). When stereotypes are activated and applied in judgment, participants render stereotype-congruent responses (e.g., selecting the "more threatening" response on trials with a Black prime). If stereotypes are activated but not applied, participants render stereotype-incongruent responses (e.g., selecting the "less threatening" response on trials with a Black prime). Validation of the SMT process model showed that modeling the absence of stereotype application as a process of contrast against activated stereotypes best accounted for experimental data (Krieglmeyer & Sherman, 2012). If stereotypes are not activated (1 − SAC), participants may correctly detect (D) the threat level of the target (e.g., selecting the "more threatening" response on trials with a high threat target). However, if detection fails (1 − D), participants guess "more threatening" with the probability (G) or "less threatening" with the probability of (1 − G).

The four processes of the SMT model are estimated using the frequencies of "more threatening" and "less threatening" responses given for the different prime and target combinations. Each branch of the SMT model represents the joint influences of cognitive processes that lead to "more" and "less" responses. Probability estimates for the branches are a product of the probabilities of processes that make up the branch. For example, on a trial with a Black prime and low threat target, a (incorrect) "more threatening" response may result from stereotype activation and stereotype application (SAC × SAP). Alternatively, the same "more threatening" response could also result from a guessing bias when the stereotype is not active and the target threat is not correctly detected ([1 − SAC] × [1 − D] × G). As such, when a trial has a Black prime and a low threat target, the probability of a high threat judgment is ([SAC × SAP] + [1 − SAC] × [1 − D] × G).

All of the parameters are estimated using an expectation minimization algorithm to arrive at a maximum likelihood solution, in which the model expectations best approximate observed data. We estimated all models using the freely available computer program MultiTree (Moshagen, 2010). The degree to which there is a discrepancy between model expectations and observed data is reflected in the G^2 goodness of fit test. When the model's expectations closely approximate the observed data, it is reflected in small G^2 and high p values. The same goodness of fit tests assess whether parameter estimates are statistically different across manipulations. Here, large G^2 and low p values indicate that parameters significantly differ from each other.

Multinomial modeling analyses. We aggregated "more threatening" and "less threatening" responses for each SMT trial type, and fit the SMT model to the Safe, Accurate, and Quick intention conditions. The model fit appeared to be somewhat poor, $G^2(6) = 43.83$, $p < .001$. However, G^2 is sensitive to sample size (in this case, fit to $N = 30{,}240$ responses) and can detect even small misfit in large samples (Cressie, Pardo, & del Carmen Pardo, 2003). The *Phi* coefficient assesses the magnitude, or effect size, of model misfit. The resulting estimate, $\phi = .038$, indicated that the magnitude of misfit fell below Cohen's (1992) criteria for a "small" effect ($\phi = .10$) after controlling for statistical power.

To examine what processes could account for the effectiveness of implementation intentions, we fit the SMT model to data from the three conditions: think "safe" versus "accurate" versus "quick." We first fit a baseline model that allowed all parameters to vary freely. Then, to test for differences between conditions in parameter estimates, we constrained each parameter one at a time across conditions. A significant reduction in model fit from the baseline model suggests that the parameter estimates differ reliably across intention conditions.

First, we investigated whether the SAC parameter differed across intention conditions.[7] Safe intentions reduced stereotype activation compared to Accurate intentions, $\Delta G^2(1) = 64.56$, $p < .001$, $w = .05$ (see Table 2 for model estimates), and compared to Quick intentions, $\Delta G^2 = 101.00$, $p < .001$, $w = .06$. This indicates that the "think safe" intentions reduced stereotype activation relative to both control conditions. Likewise, "think safe" intentions reduced stereotype application relative to both the "think accurate" control, $\Delta G^2(1) = 252.92$, $p < .001$, $w = .09$, and the "think quick" control, $\Delta G^2 (1) = 445.57$, $p < .001$, $w = .12$. There were no differences between conditions on the D parameter[8] (all $ps > .686$). There was a greater rate of guessing "more threatening" in the Safe condition than the Accurate, $\Delta G^2(1) = 48.46$, $p < .001$, $w = .04$, and Quick conditions, $\Delta G^2(1) = 53.27$, $p < .001$, $w = .04$. Such a guessing bias cannot account for the effects of implementation intentions on the SMT effect, as increased threat judgments in the Safe condition would not reduce the extent of racial bias.[9]

Discussion

Replicating prior research using implementation intentions, safe implementation intentions significantly reduced stereotyping, as measured by the SMT. Specifically, safe intentions decreased the proportion of threat judgments given to targets following Black primes relative to White primes. This effect coincided with reductions to both stereotype activation and application. Implementation intentions did not appear to have any effect on detection, suggesting that these particular implementation intentions do not affect the ability to detect target threat level.

Experiment 2

The results of Experiment 1 indicate that forming implementation intentions to "think safe" in response to Black faces reduced stereotyping on the SMT, the extent of stereotype activation, and the extent of stereotype application. Given that these effects occurred within the context of an implicit

Table 2. SMT Model Parameter Estimates [and 95% Confidence Intervals] by Implementation Intention Condition.

Experiment 1			
	Accurate	Quick	Safe
SAC	.55 [.49, .61]	.63 [.58, .70]	.22 [.17, .28]
SAP	.62 [.59, .64]	.61 [.59, .64]	.00 [–.12, .12]
D	.06 [.04, .09]	.07 [.04, .09]	.06 [.04, .08]
G	.26 [.24, .27]	.25 [.23, .27]	.33 [.32, .34]

Note. SMT = stereotype misperception task; SAC = stereotype activation; SAP = stereotype application; D = target detection; G = guessing.

measure of bias, implementation intentions are certainly relatively efficient. In addition, implementation intentions impacted judgments despite participants being explicitly asked to not be influenced by the prime stimuli, suggesting that the implementation intentions were effective even when participants intended to avoid the influence of the primes. Nevertheless, it is possible that using implementation intentions does require a modicum of cognitive resources. In the remainder of our experiments, we directly investigate the extent to which implementation intentions rely on cognitive resources to reduce implicit bias. To our knowledge, no prior research has investigated how cognitive resource restrictions influence the effectiveness of implicit bias interventions. Establishing whether implementation intentions can efficiently reduce bias in an intergroup context is important, as prior research has noted that stereotyping is more likely when individuals have limited cognitive resources (Gilbert & Hixon, 1991). Furthermore, previous work has found that interacting with outgroup members can be cognitively loading (Richeson & Shelton, 2003). While examining the efficiency of implementation intentions for reducing implicit bias, we also examine how these conditions affect the specific processes through which implementation intentions may reduce bias, namely, stereotype activation and application. In Experiment 2, we investigate the role of cognitive resources via the use of a dual-task demand to vary cognitive load.

Method

Participants. In Experiment 2, we planned to collect at least 82 participants to be powered at 95%, based on the effect size of $\eta_p^2 = .09$ reported in Stewart and Payne (2008). In total, 117 participants from the University of California, Davis completed the experiment for partial course credit. Two participants made the same judgment on every trial, and an additional five participants failed to correctly report their condition assignment. The final sample size was 110 ("safe" $N = 58$, "quick" $N = 52$).

Stimuli, procedure, and design. Prime and target stimuli were identical to the stimuli used in Experiment 1. Experiment 2 introduced a factor of cognitive load via a digit-span manipulation. On one block of SMT trials, participants were asked to keep a nine-digit number in memory (high load), whereas on the other block, participants were asked to keep a two-digit number in memory (low load). Because the Accurate and Quick conditions did not differ in Experiment 1, we dropped the Accurate condition in the experiments that followed (also see Stewart & Payne, 2008).

This resulted in a 2 (Implementation Intention: Quick vs. Safe) × 2 (Cognitive Load: High vs. Low) × 2 (Load order: High to Low vs. Low to High) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target Threat: High vs. Low) mixed design. Implementation intentions and load order were manipulated between-subjects, and all other factors were within-subjects. Load order was counterbalanced. The order of cognitive load had no effect on the comparisons of interest and, as such, will not be mentioned further.

Results

SMT effect. To examine the effects of cognitive load on implementation intentions, we subjected the proportion of "more threatening" judgments to a 2 (Intention: Quick vs. Safe) × 2 (Cognitive Load: Low vs. High) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target Threat: High vs. Low) mixed ANOVA. There was a main effect of Target, $F(1, 108) = 35.52$, $p < .001$, $\eta_p^2 = .25$, such that high threat targets received a greater proportion of threat judgments than low threat targets (see Table 3). There also was a prime main effect, $F(2.34, 205.76) = 3.34$, $p = .030$, $\eta_p^2 = .03$. Simple comparisons indicated that targets following White primes received a greater proportion of threat judgments than targets following Neutral primes, $F(1, 109) = 9.16$, $p = .003$, $\eta p^2 = .08$. There was no difference in the proportion of threat judgments following Black primes relative to White ($p = .098$) or Neutral ($p = .594$) primes. Finally, there was a main effect of intention, $F(1, 108) = 6.52$, $p = .012$, $\eta_p^2 = .06$, such that a greater proportion of threat judgments were given in the Quick than in the Safe condition.

If cognitive load reduces the effectiveness of implementation intentions, we would expect an interaction among load, implementation intentions, and prime. However, such an interaction with load was not observed, suggesting that cognitive resource restrictions had no detectable impact on the effectiveness of implementation intentions, $F(1.91, 206.16) = 2.07$, $p = .131$, $\eta_p^2 = .02$. The results replicated the interaction between intention and prime observed in Experiment 1, $F(1.91, 205.76) = 14.92$, $p < .001$, $\eta_p^2 = .12$ (see Figure 4). To better understand this interaction, we examined the prime effect within both the Quick and Safe conditions. Within the Quick condition, there was a main effect of prime, $F(2, 102) = 6.79$, $p = .002$, $\eta_p^2 = .12$. Simple comparisons showed that the proportion of threat judgments was higher following Black primes than White primes, $F(1, 51) = 5.00$, $p = .030$, $\eta_p^2 = .09$, 95% CIdiff [.01, .20], and Neutral primes, $F(1, 51) = 12.19$, $p = .001$, $\eta_p^2 = .19$, 95% CIdiff [.07, .26].

Table 3. Proportion of Threat Judgments (and Standard Errors) as a Function of Implementation Intention, Cognitive Load, Prime, and Target.

Experiment 2						
	Low load			High load		
	Black	Neutral	White	Black	Neutral	White
Quick						
High threat	.49 (.04)	.36 (.04)	.38 (.04)	.50 (.03)	.38 (.04)	.41 (.04)
Low threat	.45 (.03)	.25 (.04)	.35 (.04)	.48 (.03)	.28 (.04)	.35 (.04)
Safe						
High threat	.23 (.03)	.33 (.04)	.43 (.03)	.21 (.03)	.40 (.04)	.46 (.03)
Low threat	.21 (.03)	.25 (.04)	.40 (.04)	.19 (.03)	.29 (.04)	.41 (.03)

Note. Primes listed horizontally in the top row. Targets listed vertically in column.

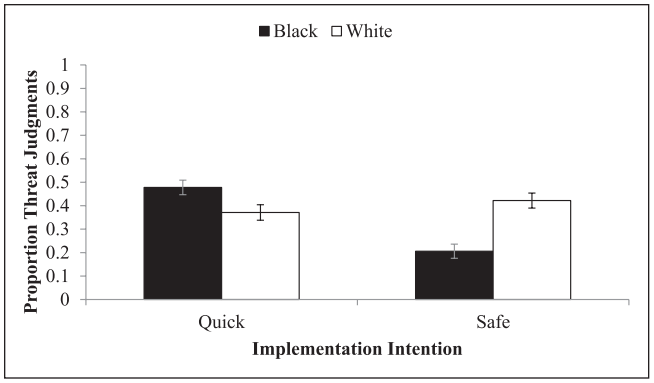

Figure 4. Proportion of threat judgments as a function of implementation intention and prime in Experiment 2.
Note. Error bars represent standard error of the mean.

Table 4. SMT Model Parameter Estimates [and 95% Confidence Intervals] by Implementation Intention.

Experiment 2		
	Quick	Safe
SAC	.62 [.56, .67]	.21 [.56, .60]
SAP	.58 [.56, .60]	.00 [.09, .09]
D	.10 [.07, .12]	.06 [.05, .08]
G	.29 [.28, .31]	.27 [.26, .29]

Note. SMT = stereotype misperception task; SAC = stereotype activation; SAP = stereotype application; D = target detection; G = guessing.

In the Safe condition, there also was a prime main effect, $F(1.90, 105.62) = 11.60$, $p < .001$, $\eta_p^2 = .17$. Simple comparisons revealed a lower proportion of threat judgments following Black primes than White primes, $F(1, 57) = 18.90$, $p < .001$, $\eta_p^2 = .25$, 95% CIdiff [−.31, −.12], and Neutral primes, $F(1, 57) = 5.57$, $p = .022$, $\eta_p^2 = .09$, 95% CIdiff [−.20, −.02]. Once again, the Safe instructions resulted in a reversal of the standard SMT stereotyping effect. In addition, a greater proportion of threat judgments were given following White primes than Neutral primes, $F(1, 57) = 8.16$, $p = .006$, $\eta_p^2 = .13$, 95% CIdiff [.03, .18].

Multinomial modeling analyses. We fit the SMT model to the Safe and Quick intention conditions. As in Experiment 1, there was a detectable discrepancy between model predictions and the observed data. However, after controlling for power, the magnitude of misfit was small, $G^2 = 57.19$, $p < .001$, $\phi = .043$. Replicating Experiment 1, SAC was reduced in the Safe relative to the Quick condition, $\Delta G^2 = 111.47$, $p < .001$, $w = .06$ (see Table 4 for parameter estimates). Likewise, SAP was reduced when participants formed intentions to think Safe versus Quick, $\Delta G^2 = 309.94$, $p < .001$, $w = .10$.

Discussion

Experiment 2 replicated Experiment 1 in finding that Safe implementation intentions reduced stereotyping relative to Quick intentions. As in Experiment 1, this effect was associated with reductions in both stereotype activation and stereotype application. Cognitive load had no detectable effect on the effectiveness of implementation intentions in reducing SMT stereotyping, stereotype activation, or stereotype application.

Experiment 3

Experiment 2 found that the effectiveness of implementation intentions to "think safe" in reducing stereotyping was not affected by the availability of cognitive resources. In addition, Experiment 2 provided further evidence that implementation intentions reduce both stereotype activation and stereotype application.

In Experiment 3, we sought to test whether implementation intentions would prove effective at stereotype reduction when cognitive resources were constrained by the time participants were given to respond. Likewise, we sought to determine whether implementation intentions would continue to reduce both stereotype activation and stereotype application when time was limited.

Table 5. Proportion of Threat Judgments (and Standard Errors) as a Function of Response Window, Implementation Intention, Prime, and Target.

Experiment 3

	Quick			Safe		
	Black	Neutral	White	Black	Neutral	White
Long window						
High threat	.49 (.03)	.30 (.04)	.31 (.03)	.21 (.03)	.37 (.04)	.41 (.03)
Low threat	.47 (.03)	.23 (.04)	.31 (.03)	.19 (.03)	.33 (.04)	.43 (.03)
Short window						
High threat	.47 (.03)	.37 (.04)	.40 (.03)	.27 (.03)	.33 (.04)	.32 (.03)
Low threat	.49 (.03)	.38 (.04)	.37 (.03)	.25 (.03)	.31 (.04)	.33 (.03)

Note. Primes listed horizontally in the top row. Targets listed vertically in column.

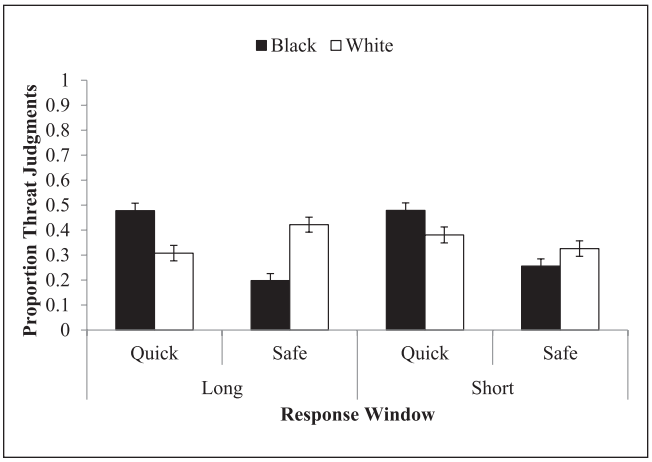

Figure 5. Proportion of threat judgments as a function of response window, implementation intention, and prime in Experiment 3.
Note. Error bars represent standard errors of the mean.

Method

Participants. In Experiment 3, we planned to sample from at least 192 participants to be powered at 80%, based on the effect size, Cohen's $f = .29$, found in Experiment 1. In total, 201 participants from the University of California, Davis completed the experiment for partial course credit. Four participants made the same judgment on every trial, one participant was identified as an outlier according to our preregistered plan, and an additional 20 participants failed to accurately report their condition assignment. The final sample size was 175 participants ("safe" $N = 93$, "quick" $N = 85$).

Stimuli, procedure, and design. The prime and target stimuli were identical to the stimuli used in Experiments 1 and 2. Experiment 3 introduced a between-subjects manipulation of response window. Participants were randomly assigned to either a version of the SMT in which they were required to respond within 450 ms (short) or a version in which they were required to respond within 850 ms (long). Our selection of response windows was determined by the approximate time participants usually take to respond on the SMT (Studies 1 and 2 $M_{RT} = 594.53$) and corresponds to common conventions for speeded and nonspeeded responding in sequential priming paradigms, such as the weapon identification task (e.g., Payne, 2001). This resulted in a 2 (Implementation Intention: Safe vs. Quick) × 2 (Response Window: Short vs. Long) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target Threat: High vs. Low) mixed design, in which implementation intentions and response window were between-subjects, and all other factors were within-subjects.

Results

SMT effect. To examine whether response window moderated the effect of implementation intentions, we subjected the proportion of "more threatening" judgments to a 2 (Intention: Quick vs. Safe) × 2 (Response Window: Short vs. Long) × 3 (Prime: Black vs. Neutral vs. White) × 2 (Target Threat: High vs. Low) mixed ANOVA. There was a Target main effect, $F(1, 174) = 4.60$, $p = .033$, $\eta_p^2 = .03$, indicating that high threat targets received a greater proportion of threat judgments than low threat targets (see Table 5). There also was a main effect of intention, $F(1, 174) = 7.96$, $p = .005$, $\eta_p^2 = .04$, such that there was a greater proportion of threat responses given in the Quick than in the Safe condition.

Replicating the prior two experiments, there was an Intention × Prime interaction, $F(2, 348) = 37.00$, $p < .001$, $\eta_p^2 = .18$. However, this effect was qualified by a three-way interaction among intention, response window, and prime, $F(2, 348) = 5.76$, $p = .003$, $\eta_p^2 = .03$ (see Figure 5). To better understand this interaction, we conducted two separate Intention × Prime ANOVAs on the proportion of "more threatening" judgments, one for each level of the response window manipulation.

In the Long Window condition, there was an Intention × Prime interaction, $F(2, 178) = 29.94$, $p < .001$, $\eta_p^2 = .25$. As in Experiments 1 and 2, the Quick condition showed a

Table 6. SMT Model Parameter Estimates [and 95% Confidence Intervals] by Implementation Intention.

Experiment 3				
	Long		Short	
	Quick	Safe	Quick	Safe
SAC	.55 [.48, .62]	.21 [.16, .25]	.43 [.27, .60]	.06 [.02, .10]
SAP	.66 [.62, .69]	.00 [–.12, .11]	.61 [.56, .67]	.00 [–.37, .37]
D	.05 [.02, .09]	.03 [.01, .06]	.00 [–.03, .03]	.01 [–.01, .03]
G	.25 [.23, .27]	.29 [.28, .31]	.38 [.36, .40]	.29 [.28, .30]

Note. SMT = stereotype misperception task; SAC = stereotype activation; SAP = stereotype application; D = target detection; G = guessing.

standard prime main effect, $F(2, 84) = 17.00$, $p < .001$, $\eta_p^2 = .29$. Simple comparisons showed that the proportion of threat judgments was greater following Black primes than White primes, $F(1, 42) = 13.79$, $p = .001$, $\eta_p^2 = .25$, 95% CIdiff [.08, .26], and Neutral primes, $F(1, 42) = 30.01$, $p < .001$, $\eta_p^2 = .42$, 95% CIdiff [.13, .29]. Once again, the Safe condition showed a reversed prime effect, $F(2, 94) = 15.52$, $p < .001$, $\eta_p^2 = .25$, with a lower proportion of threat judgments following Black primes than White primes, $F(1, 47) = 29.60$, $p < .001$, $\eta_p^2 = .39$, 95% CIdiff [−.31, −.14], and Neutral primes $F(1, 47) = 12.32$, $p = .001$, $\eta_p^2 = .21$, 95% CIdiff [−.24, −.06].

In the Short Window condition, there also was an Intention × Prime interaction, $F(2, 170) = 8.73$, $p < .001$, $\eta_p^2 = .09$. The Quick condition showed a standard prime effect, $F(2, 82) = 7.31$, $p = .001$, $\eta_p^2 = .15$. Simple comparisons showed that a greater proportion of threat judgments were made following Black primes than White primes, $F(1, 41) = 9.71$, $p = .003$, $\eta_p^2 = .19$, 95% CIdiff [.04, .16], and Neutral primes, $F(1, 41) = 9.00$, $p = .005$, $\eta_p^2 = .18$, 95% CIdiff [.03, .17]. In the Safe condition, there was a trend toward a reversed prime effect, $F(2, 88) = 2.49$, $p = .088$, $\eta_p^2 = .05$. Simple comparisons showed that the proportion of threat judgments was lower following Black primes than White primes, $F(1, 44) = 4.01$, $p = .051$, $\eta_p^2 = .08$, 95% CIdiff [−.14, .00], but not Neutral primes, $F(1, 44) = 2.74$, $p = .105$. Thus, we observed the same interaction between prime and intention as in the previous studies, although the interaction was weaker when the response window was shorter.

Multinomial modeling analyses. The extent of model misfit was small in magnitude after controlling for power, $G^2(8) = 67.19$, $p < .001$, $\phi = .051$. To examine the interaction between response window and implementation intentions, we estimated a model that permitted an interaction between intentions and response window, and compared its fit with a model that did not permit the interaction. Analyses revealed an interaction between intention and response window on the SAC parameter, $\Delta G^2(1) = 6.99$, $p = .008$, $w = .02$. Forming intentions to "think safe" reduced SAC at both Short and Long windows. However, the reduction in SAC was larger in the Long condition, $\Delta G^2(1) = 51.72$, $p < .001$, $w = .04$, than in the Short condition, $\Delta G^2(1) = 14.33$, $p < .001$, $w = .02$ (see Table 6). Replicating prior experiments, forming intentions to "think safe" reduced SAP, $\Delta G^2(1) = 254.88$, $p < .001$, $w = .10$. This effect was not moderated by response window, $G^2(1) < .001$, $p > .999$, $w < .001$.

Discussion

Consistent with Experiments 1 and 2, forming intentions to "think safe" reduced racial stereotyping on the SMT. Likewise, implementation intentions reduced both stereotype activation and stereotype application. Unlike in Experiment 2, we found that restricting cognitive resources by restricting the time participants had to reach their judgments moderated the effectiveness of implementation intentions. Specifically, implementation intentions were less effective in reducing stereotyping when the response window was short compared to when the response window was longer. However, implementation intentions were still effective in reducing implicit stereotyping, even when participants had to respond within the short response window. Likewise, the reduction in stereotype activation due to safe intentions was present under both response windows but reduced in the short window. These findings suggest that intentions operate efficiently, but not without limit; their effectiveness may be reduced when people must rapidly make judgments. The extent to which safe intentions reduced stereotype application was not affected by response window.

Experiment 4

In Experiment 4, we sought to replicate and extend our previous results by testing stereotyping of a different social category: gender. Rather than using primes that varied in race, we used male and female primes. In doing so, we can gain further insight into the conditions under which implementation intentions are effective. Much research has observed that participants from college undergraduate samples have become increasingly hesitant to endorse racial stereotyping (Gaertner & Dovidio, 1986; Schuman, Steeh, Bobo, & Krysan, 1997). In some cases, students hold strong egalitarian goals that allow them to regulate their bias, even on implicit measures (Moskowitz & Li, 2011). As such, we assume that

Table 7. Proportion of Threat Judgments (and Standard Errors) as a Function of Implementation Intention, Response Window, Prime, and Target.

Experiment 4

	Safe			Quick		
	Black	Neutral	White	Black	Neutral	White
Long window						
High threat	.33 (.03)	.26 (.04)	.31 (.03)	.60 (.03)	.42 (.04)	.24 (.03)
Low threat	.32 (.03)	.23 (.04)	.31 (.03)	.58 (.03)	.41 (.03)	.26 (.03)
Short window						
High threat	.28 (.03)	.26 (.04)	.36 (.03)	.51 (.03)	.41 (.04)	.32 (.03)
Low threat	.30 (.03)	.26 (.04)	.25 (.03)	.50 (.03)	.40 (.03)	.31 (.03)

Note. Primes listed horizontally in the top row. Targets listed vertically in column.

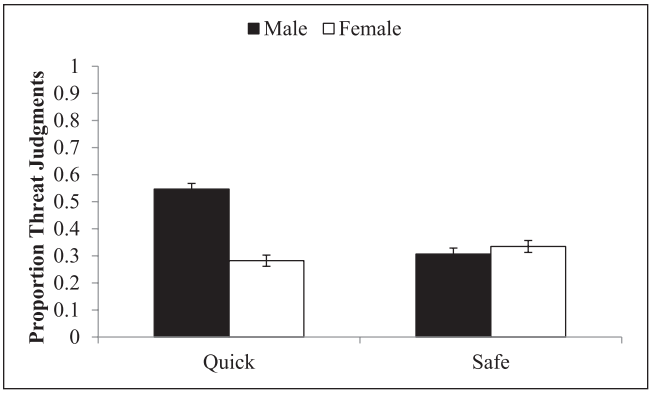

Figure 6. Proportion of threat judgments as a function of implementation intention and prime in Experiment 4.
Note. Error bars represent standard errors of the mean.

many of our participants possess a goal to avoid using racial stereotypes and have some practice with doing so. In contrast, although men are generally stereotyped as more aggressive and threatening than women (Swim, 1994; Weaver, Vandello, Bosson, & Burnaford, 2010), it does not seem that people spontaneously correct for a bias to judge men as more threatening than women. As such, we expected our participants to have less practice with and less motivation to avoid judging men as more threatening than women. Thus, investigating gender rather than race allows us to determine whether implementation intentions can reduce stereotyping even when participants have little practice with and motivation to correct for the stereotype. Furthermore, in manipulating response window, we can once again test the extent to which intentions operate efficiently, but now for a goal that participants are unlikely to hold intrinsically and for which they are unlikely to be practiced.

Participants

In Experiment 4, we planned to sample from at least 192 participants to be powered at 80%, based on the effect size, Cohen's $f = .29$, found in Experiment 1. In total, 231 participants from the University of California, Davis completed the experiment for partial course credit. A computer error resulted in the loss of data from 10 participants, and another 29 failed to accurately report their condition assignment. The final sample size was 228 participants ("safe" $N = 114$, "quick" $N = 114$).

Stimuli, Procedure, and Design

Prime stimuli were photographs of 24 men and women taken from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) that were cropped at the neck. We also continued to use the same neutral prime image from the prior experiments. All prime stimuli were superimposed on a gray background. Target stimuli were the same as used in prior experiments. The procedure was identical to Experiment 3. This resulted in a 2 (Implementation Intention: Safe vs. Quick) × 2 (Response Window: Short vs. Long) × 3 (Prime: Male vs. Neutral vs. Female) × 2 (Target Threat: High vs. Low) mixed design, in which intentions and response window were between-subjects, and other factors were within-subjects.

Results

SMT effects. We conducted an Implementation intention × Response window × Prime × Target mixed ANOVA on the proportion of "more threatening" judgments. There was a main effect of prime type, $F(1.92, 373.84) = 21.63$, $p < .001$, $\eta_p^2 = .10$. Simple comparisons showed that a greater proportion of threat judgments were given following Male primes than Female primes, $F(1, 198) = 30.00$, $p < .001$, $\eta_p^2 = .13$, and Neutral primes $F(1, 198) = 34.80$, $p < .001$, $\eta_p^2 = .15$ (see Table 7). There also was a main effect of intention, $F(1, 195) = 25.43$, $p < .001$, $\eta_p^2 = .12$, showing that a greater proportion of threat judgments were given in the Quick than the Safe condition. However, these main effects were qualified by the expected Intention × Prime interaction, $F(2, 390) = 31.29$, $p < .001$, $\eta_p^2 = .14$ (see Figure 6). Response window

Table 8. SMT Model Parameter Estimates [and 95% Confidence Intervals] by Implementation Intention.

Experiment 4		
	Quick	Safe
SAC	.26 [.42, .49]	.28 [.22, .33]
SAP	1.00 [.85, .1.15]	.45 [.42, .49]
D	.01 [–.01, .03]	.01 [–.01, .03]
G	.40 [.38, .41]	.25 [.24, .26]

Note. SMT = stereotype misperception task; CI = confidence interval; SAC = stereotype activation; SAP = stereotype application; D = target detection; G = guessing.

did not moderate the effectiveness of implementation intentions, $F(2, 390) = 0.42$, $p = .660$.

To understand the Intention × Prime interaction, we conducted two different ANOVAs examining the prime effect for each intention condition. In the Quick condition, there was a prime main effect, $F(2, 208) = 59.95$, $p < .001$, $\eta_p^2 = .37$. Simple comparisons showed that there was a greater proportion of threat judgments following Male primes than Female primes, $F(1, 104) = 116.28$, $p < .001$, $\eta_p^2 = .53$, 95% CIdiff [.22, .31], and Neutral primes, $F(1, 104) = 31.75$, $p < .001$, $\eta_p^2 = .23$, 95% CIdiff [.09, .18]. There also was a lower proportion of threat judgments following Female primes than Neutral primes, $F(1, 104) = 29.13$, $p < .001$, $\eta_p^2 = .22$, 95% CIdiff [−.18, −.08].

In the Safe condition, there was a reliable but weaker prime effect, $F(1.65, 153.10) = 3.85$, $p = .031$, $\eta_p^2 = .04$. Simple comparisons showed that the proportion of threat judgments did not reliably differ following Male and Female primes, $F(1, 93) = .61$, $p = .435$, but that there was a higher proportion of threat judgments following Male primes than Neutral primes, $F(1, 93) = 6.41$, $p = .013$, $\eta_p^2 = .06$, 95% CIdiff [.01, .10]. In addition, the proportion of threat judgments was significantly higher after Female primes than Neutral primes, $F(1, 93) = 6.24$, $p = .014$, $\eta_p^2 = .06$, 95% CIdiff [.02, .15].

Multinomial modeling analyses. We fit the SMT model to the two intention conditions. Model misfit was small after controlling for statistical power, $G^2(4) = 10.01$, $p = .040$, $\phi = .019$. Unlike Experiments 1 to 3, there was no effect of implementation intention on SAC, $\Delta G^2(1) = .26$, $p = .609$ (see Table 8). However, intentions to think Safe reduced SAP compared with intentions to think Quick, $\Delta G^2(1) = 77.25$, $p < .001$, $w = .05$.

Discussion

Experiment 4 largely replicated the previous experiments in showing that implementation intentions reduced implicit stereotyping. This effect was not moderated by response window. Although stereotyping was observed in the "think safe" condition, it was significantly reduced, extending the results of Experiments 1 to 3 in showing that implementation intentions are effective even for stereotypes that participants are unlikely to regularly inhibit. With regard to the processes underlying these effects, our findings were more mixed: Implementation intentions reduced the application of gender-based stereotypes but did not influence the activation of gender-based stereotypes.

General Discussion

The current research had two goals: to examine the processes by which implementation intentions reduce implicit bias and to test the extent to which implementation intentions reduce such bias efficiently. With respect to the first goal, results show that, in most cases (Experiments 1 to 3), implementation intentions to "think safe" reduce both stereotype activation (i.e., the extent to which stereotypic information comes to mind) and stereotype application (i.e., the extent to which stereotypes are applied in judgment). Interestingly, for gender-based stereotypes, a domain in which people have less practice and less motivation to avoid bias, implementation intentions only impacted the stereotype application process. With respect to the second goal, we found that the effect of implementation intentions on reducing implicit stereotyping was generally not diminished by constraints on cognitive resources (Experiments 2 to 4). In Experiment 3, short response windows did reduce the effectiveness of safe intentions. Nevertheless, at both response windows, safe intentions effectively eliminated typical racial biases. The cognitive processes affected by implementation intentions also were largely unaffected by resource restrictions. That is, intentions to "think safe" reduced both stereotype activation and application, regardless of cognitive resources.

Experiments 1 to 3 showed a consistent effect of safe intentions reducing race-based stereotype activation and application. In contrast, in Experiment 4, intentions to "think safe" influenced the likelihood of applying gender stereotypes but not the likelihood of activating gender stereotypes. There are a number of possible reasons why gender and race stereotyping may differ in terms of stereotype activation and why stereotyping interventions may operate differently with respect to the two groups. First, participants may be less intrinsically motivated (and, therefore, less practiced) to correct for the stereotype of males as threatening than they are to correct for the stereotype of Black people as threatening (Gaertner & Dovidio, 1986; Moskowitz & Li, 2011; Schuman et al., 1997), which could diminish people's capacity to down-regulate stereotype activation. Moreover, gender is often seen as a more essentialized category than race (Gelman, Collman, & Maccoby, 1986; Gelman & Taylor, 2000; Taylor, 1996), and gender categories are preferentially used (vs. race categories) when the two compete (e.g., Klauer, Hölzenbein, Calanchini, & Sherman, 2014). Finally, it also is possible that participants may simply have stronger mental associations of males (vs. females) with threat than of

Black males (vs. White males) with threat, which could also diminish the capacity for the down-regulation of stereotype activity. Regardless of the reasons why implementation intentions affected stereotype application but not activation of gender stereotypes, it is notable that they did successfully decrease gender stereotyping.

Implementation intentions did not have a consistent effect on target detection. Such findings have important implications for the use of implementation intentions as a bias-reduction strategy. Specifically, although implementation intentions caused people to behave in a less racially biased way, they were not necessarily more accurate in their detection of target threat. In addition, in the behavioral data, we observed reduced bias via a reversal pattern in threat judgments: Targets following White faces received a greater proportion of threat judgments than targets following Black faces. Although this is a clear reduction in Black-threat bias, such bias reduction does not appear to improve accuracy. Thus, implementation intentions are effective at decreasing racial bias but may not be an appropriate strategy for increasing accuracy in person perception. These findings emphasize the importance of better understanding how interventions influence cognitive processes and exert their effects before such strategies are applied outside of the lab.

Our second major finding was that implementation intentions were highly resistant to the restriction of cognitive resources. Intentions were effective under cognitive load, even when the intentions were formed to reduce bias toward a group that individuals likely had little intrinsic motivation to avoid stereotyping. These findings suggest that implementation intentions are highly efficient and reduce bias even when it is more difficult to keep the intentions in mind. Very little research has been conducted exploring whether bias interventions will be effective in conditions that restrict cognitive resources. For example, previous research found that intergroup interactions can be stressful, and tax executive functioning (Richeson & Shelton, 2003). Determining whether bias interventions work in nonideal situations, such as when cognitive resources are limited, is important for the application of such interventions. Future research should further test the conditions in which a variety of bias interventions can successfully operate.

Although we find consistent effects of implementation intentions for a reduction in implicit threat bias, it is unknown how this might map on to more complex situations outside of the lab. The current work focuses on the effect that implementation intentions have on mental processes that occur relatively early in person perception, wherein participants are often making their judgments in less than a second. It seems potentially promising that implementation intentions influence judgments even in cognitively taxing situations, which may more closely capture the state of perceivers in an intergroup context. In particular, interracial interaction contexts are often associated with decreased executive function for White perceivers, suggesting that intergroup interaction is a cognitively demanding experience (Richeson & Shelton, 2003). Overall, our findings likely best predict impressions and behavioral outcomes in contexts in which judgments are made quickly. It is possible that in lengthier intergroup interactions, such implementation intentions could have different effects on judgments and behavior.

Conclusion

Our research replicates and extends prior work testing the effectiveness of implementation intentions in reducing implicit bias (Lai et al., 2014). Furthering our understanding of implementation intentions, we found that safe implementation intentions reduce both stereotype activation and application but do not impact detection. These are the first data that directly examine the specific cognitive processes affected by implementation intentions. That stereotype activation was reduced may suggest that implementation intentions can temporarily change what information becomes accessible when Black faces are presented. It is not simply the case that implementation intentions affect only the application of stereotypic information that is already activated. The effects of implementation intentions on stereotype reduction processes are highly efficient and appear to not require prior practice inhibiting the stereotype in question. It will be important for future research to test whether other interventions that reduce implicit stereotyping operate similarly, specifically, whether they similarly reduce both stereotype activation and application processes or whether they alter different combinations of processes, including individuation. Our findings suggest that implementation intentions should be employed when bias reduction is the goal but not when increased judgment accuracy is the desired outcome. In addition, it is important to examine how well implicit stereotyping interventions work in cognitively restricted conditions and how motivation influences their effectiveness. A broader knowledge of the conditions under which stereotyping interventions operate, and of which processes they impact, can help us understand when and how different interventions may work most effectively.

Appendix

Additional SMT Effect Results

Experiment 1: Prime × Target Interaction.[10] There was an unpredicted Prime × Target interaction, $F(2, 207) = 12.08$, $p < .001$, $\eta_p^2 = .06$. To better understand this interaction, we examined the target effect for each prime type. These analyses revealed a target effect on trials when Black primes were presented, $F(1, 209) = 12.35$, $p = .001$, $\eta_p^2 = .06$, and when Neutral primes were presented, $F(1, 209) = 41.56$, $p < .001$, $\eta_p^2 = .17$, but not when White primes were presented, $F(1, 209) = 2.76$, $p = .098$. These results suggest that participants differentiated between target threat levels to a greater extent following Black and Neutral primes (e.g., giving a greater proportion of threat judgments to high threat targets relative to low threat targets) than they did following White primes.

Experiment 2: additional analyses. There was a Prime × Target interaction, $F(1.90, 201.10) = 17.58$, $p < .001$, $\eta_p^2 = .14$. To better understand this interaction, we examined the Target effect for each prime type. There were Target effects for all prime types: Black, $F(1, 109) = 4.40$, $p = .038$, $\eta_p^2 = .04$; Neutral, $F(1, 109) = 49.34$, $p < .001$, $\eta_p^2 = .31$; and White, $F(1, 109) = 14.90$, $p < .001$, $\eta_p^2 = .12$. In every case, high threat targets were given a greater proportion of threat judgments than low threat targets. However, the target effect was strongest on Neutral prime trials.

Experiment 3:Window × Prime × Target interaction. There was a Window × Prime × Target interaction, $F(2, 174) = 3.28$, $p = .040$, $\eta_p^2 = .02$. To better understand this interaction, we conducted separate Prime × Target ANOVAs for each Window condition. Within the Long window, there was a Prime × Target interaction, $F(2, 178) = 6.25$, $p = .002$, $\eta_p^2 = .07$, but not in the Short window condition, $F(2, 170) = .04$, $p = .963$. At Long windows, there was a target effect for Neutral primes only, $F(1, 90) = 15.00$, $p < .001$, $\eta_p^2 = .14$, in which high threat targets received a greater proportion of threat responses than low threat targets.

Experiment 4:Window × Prime interaction. There was a Response window × Prime interaction, $F(2, 390) = 4.42$, $p = .013$, $\eta_p^2 = .02$. To decompose the interaction, we conducted two separate within-subjects ANOVAs for each Window condition. In the Long window condition there was a prime main effect, $F(2.00, 188.00) = 16.45$, $p < .001$, $\eta_p^2 = .14$. Simple comparisons indicated that targets following Male primes were given a greater proportion of threat responses than Female, $F(1, 100) = 26.44$, $p < .001$, $\eta_p^2 = .21$, or Neutral primes $F(1, 100) = 21.00$, $p < .001$, $\eta_p^{2p} = .17$. In the Short window condition, there was also a prime main effect, $F(1.63, 158.00) = 5.45$, $p = .009$, $\eta_p^2 = .05$. Simple comparisons indicated that targets following Male primes were given a greater proportion of threat responses than Female, $F(1, 97) = 6.00$, $p = .017$, $\eta_p^2 = .06$, or Neutral primes $F(1, 97) = 16.00$, $p < .001$, $\eta_p^2 = .14$. Although there was a prime main effect at both windows, it appeared that the prime effect was stronger in the Long response window condition than Short.

Detection and Guessing Results by Experiment

Experiment 2. There was a significant difference between the Safe and Quick conditions on the G parameter, $\Delta G^2 = 4.60$, $p = .032$, $w = .01$, indicating that participants in the Quick condition had a tendency to guess "more threat" on a greater proportion of trials than participants in the Safe condition. The D parameter was also higher in the Quick condition than the Safe condition, $\Delta G^2 = 5.65$, $p = .020$, $w = .01$.

Experiment 3. There was an interaction between intention and response window on the G parameter, $\Delta G^2(1) = 45.16$, $p < .001$, $w = .04$. In the Long window condition, participants in the Safe condition guessed "more threatening" on a greater proportion of trials than in the Quick condition, $\Delta G^2(1) = 8.66$, $p = .003$, $w = .02$. These effects were reversed in the Short Window condition: Participants in the Quick condition guessed "more threatening" on a greater proportion of trials than in the Safe condition, $\Delta G^2(1) = 47.87$, $p < .001$, $w = .04$. Our guessing findings once again differ from the prior two experiments, making it difficult to interpret the effect of implementation intentions on guessing. Response window did not interact with the effect of intention on detection. However, there was an unsurprising effect of response window on the D parameter $\Delta G^2(1) = 5.56$, $p = .018$, $w = .01$, indicating that detection was higher in the long than short window condition.

Experiment 4. There was an effect of intentions on the G parameter, $\Delta G^2(1) = 254.96$, $p < .001$, $w = .09$, indicating that participants in the Quick condition guessed "more threatening" on a greater proportion of trials than in the Safe condition. These findings were consistent with Experiment 2. However, given the lack of consistency in effects on the G parameter, we hesitate to interpret such results. In this experiment, there was no effect of either intentions or response window on the D parameter.

AIC/BIC/MDL Information Criteria by Experiment and Multinomial Model.

	Experiment 1			Experiment 2			Experiment 3			Experiment 4		
	AIC	BIC	MDL	AIC	BIC	MDL	AIC	BIC	MDL	AIC	BIC	MDL
SMT	39,632.7	39,666	19,829	40,645.6	40,679	20,335.6	33,364	33,396.7	16,694.4	36,756.9	36,790	18,391
d-SMT	39,649.4	39,682.7	19,836.9	40,664.5	40,698	20,344.5	33,364.7	33,397.4	16,694.3	36,757.1	33,397.4	18,390.6
AMP	39,626.4	39,668	19,828.9	40,650.7	40,692.5	20,341.1	33,363	33,403.9	16,696.8	36,759	36,800.3	18,395
C-PDP	46,317.7	46,334.3	23,166.4	47,626.4	47,643.1	23,820.8	39,320.8	39,337.2	19,667.9	42,615.2	42,631.7	21,315.1
C-PDPg	39,835.7	39,860.7	19,928.1	40,836.8	40,861.9	20,428.7	36,760.9	36,785.7	16,702.1	36,760.9	36,785.7	18,390.6
A-PDP	42,360.9	42,377.5	21,188.1	43,729.6	43,746.4	21,872.5	36,132.1	36,148.4	18,073.6	39,209.6	39,266.1	19,612.4
A-PDPg	39,835.7	39,860.7	19,928.2	40,836.8	40,861.9	20,428.8	33,384.2	33,408.7	16,702.2	36,761	36,785.8	18,390.7

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; MDL = minimum description length; Model acronyms; SMT = stereotype misperception; d-SMT = detection-first SMT; AMP = affect misattribution; C-PDP = control-dominant process dissociation; C-PDPg = control-dominant PDP with guessing parameter; A-PDP = automaticity-dominant process dissociation; A-PDPg = automaticity-dominant PDP with guessing parameter.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this article was partially supported by an Anneliese Maier Research Award from the Alexander von Humboldt Foundation to Jeffrey W. Sherman.

Notes

1. In this paper, we use the term *implicit bias* to refer to bias that is measured indirectly using implicit measures (Brannon & Gawronski, 2017). Implicit measures use task performance (e.g., error rates) to infer bias rather than self-reported beliefs. This definition does not make assumptions about the nature of underlying mental processes and representations, which must be investigated empirically.
2. Although prior research using such implementation intentions for implicit bias reduction (Stewart & Payne, 2008) does not report removing participants for failing an attention check, we felt that it was sensible to remove participants who were unable to report what condition they were in, as this indicates that the intended processing goal was not effectively manipulated among these participants.
3. As our criteria for removing these participants (subjective reports from research assistants) were not mentioned in the preregistration, we conducted our analyses with the two participants in the dataset and with them removed. We found that our statistical conclusions did not differ based on whether these participants were included. Statistics reported in the main text exclude these two cases, as we believe they offer the most informational value.
4. We report all measures, conditions, data exclusions, and sample size determination criteria.
5. Huynh–Feldt corrections applied for violations of sphericity.
6. These effects were qualified by an unpredicted Prime × Target interaction, $F(2, 207) = 12.08$, $p < .001$, $\eta_p^2 = .06$. Because the prime by target interaction was unpredicted, not consistently observed across experiments, and irrelevant to the effect of implementation intentions, we fully describe this interaction in the appendix.
7. As illustrated in Figure 3, the stereotype misperception task (SMT) model is fit to response patterns for each combination of prime and target. Because the factors of prime and target are necessarily already accounted for in the model, only comparisons between additional conditions are meaningful (in this case intention type)
8. The D parameter can be conceptualized as the extent to which participants were accurate at detecting threat level on the SMT. In general, accuracy is quite low in the SMT, as target ambiguity is high. This is necessary to observe the effects of the primes. It is notable that the implementation intention manipulation to think "accurate" did not increase detection for participants. Similar results were observed by Stewart and Payne (2008). What is most important for this control condition is that participants were thinking about being accurate when they saw Black faces, whether that affected their responses or not.
9. The effects of implementation intentions on D and G parameters were not consistent across experiments and were not of primary interest. As such, we report the D and G results for subsequent studies in the appendix.
10. The prime by target interaction trended toward violating the assumption of sphericity ($p = .064$); however, because this violation was marginal, and the correction had no substantive effect on the effects, we report uncorrected statistics.

Supplemental Material

Supplementary material is available online with this article.

References

Banaji, M. R., & Greenwald, A. G. (1994). *Implicit stereotyping and prejudice*. In M. P. Zanna & J. M. Olson (Eds.), The psychology of prejudice: The Ontario symposium (pp. 55-76). Mahwah, NJ: Psychology Press.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology*, *81*, 946-960.

Brannon, S. M., & Gawronski, B. (2017). A second chance for first impressions? Exploring the context-(in) dependent updating of implicit evaluations. *Social Psychological & Personality Science*, *8*, 275-283.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469-487.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314-1329.

Cressie, N., Pardo, L., & del Carmen Pardo, M. (2003). Size and power considerations for testing loglinear models using ϕ-divergence test statistics. *Statistica Sinica*, *13*, 555-570.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800-814.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 339-360). New York, NY: Guilford Press.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 137-163). Mahwah, NJ: Psychology Press.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013-1027.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*, 307-316.

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single reading of a short scenario. *Psychonomic Bulletin & Review*, *12*, 139-144.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). San Diego, CA: Academic Press.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *57*, 396-404.

Gelman, S. A., & Taylor, M. G. (2000). Gender essentialism in cognitive development. In P. H. Miller & E. Kofsky Scholnick (Eds.), *Toward a feminist developmental psychology* (pp. 169-190). Florence, KY: Taylor & Frances/Routledge.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509-517.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, *38*, 69-119.

Govorun, O., & Payne, B. K. (2006). Ego—Depletion and prejudice: Separating automatic and controlled components. *Social Cognition*, *24*, 111-136.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.

Hehman, E., Flake, J. K., & Calanchini, J. (2017). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*. Advance online publication. doi:10.1177/1948550617711229

Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology*, *43*, 137-146.

Klauer, K. C., Hölzenbein, F., Calanchini, J., & Sherman, J. W. (2014). How malleable is categorization by race? Evidence for competitive category use in social categorization. *Journal of Personality and Social Psychology*, *107*, 21-40.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, *103*, 205-224.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*, 522-544.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E., Joy-Gaba, J. A., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765-1785.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York, NY: Guilford Press.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122-1135.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*, 512-523.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42-54.

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*, 103-116.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087-11092.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181-192.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and Blacks with approach behaviors. *Journal of Personality and Social Psychology*, *100*, 197-210.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287-290.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America*. Cambridge, MA: Harvard University Press.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*, 1332-1345.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21-36.

Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, *67*, 1555-1571.

Weaver, J. R., Vandello, J. A., Bosson, J. K., & Burnaford, R. M. (2010). The proof is in the punch: Gender differences in perceptions of action and aggression as components of manhood. *Sex Roles*, *62*, 241-251.

Webb, T. L., Sheeran, P., & Pepper, J. (2012). Gaining control over responses to implicit attitude tests: Implementation intentions engender fast responses on attitude-incongruent trials. *British Journal of Social Psychology*, *51*, 13-32.